

DOCUMENT RESUME

ED 088 897

TM 003 406

AUTHOR Mason, Ward S.
TITLE A Critique of the Rite of Passage Review of Laboratory and Center Programs Prepared for the Mid Course Assessment.
INSTITUTION National Inst. of Education (DHEW), Washington, D.C. Office of Research and Development Resources.
PUB DATE 16 May 72
NOTE 14p.
EDRS PRICE MF-\$0.75 HC-\$1.50
DESCRIPTORS Decision Making; Evaluation Methods; Performance Criteria; *Performance Specifications; Program Effectiveness; *Program Evaluation; *Regional Laboratories; *Research and Development Centers
IDENTIFIERS *National Institute of Education; NIE; NIE Archives; Program Review; Rite of Passage Review

ABSTRACT

This critique lists a number of positive features of the Rite of Passage Review and dwells at somewhat more length on its problems and difficulties. The remarks are organized around several key features of the design which were taken to operationalize the process: 1) Specialist Panel Personnel; 2) Size of Panels; 3) Orientation of Panel Members and Chairmen; 4) Cross-Institutional Substantive Review of Programs; 5) Review Methodology; and, 6) Problems of Consistency Within and Across Panels. (RF)

TM

A CRITIQUE OF THE RITE OF PASSAGE REVIEW OF
LABORATORY AND CENTER PROGRAMS PREPARED FOR THE
MID COURSE ASSESSMENT a/

Ward S. Mason
Acting Assistant Director
Division of Research and Development Resources

The following observations of the Rite of Passage Review are offered in the spirit of constructive criticism in the hope that the review itself and the later development of the complete assessment system will be of very high quality. We espouse the importance of evaluation in government programs, and it is essential that we evaluate our own efforts.

It should be understood that this review is taking place in a context which was not anticipated. The plans for the review were drawn up on the assumption that NIE would have become a reality in the Fall of 1971, that the Master Panel would have been appointed first, the MP would have spent two months in training, would have helped select and direct the Specialist Panels, etc., etc. Clearly this is not the context in which we are operating.

The critique below lists a number of positive features of the Rite of Passage Review and dwells at somewhat more length on problems and difficulties - as is typical of critiques and site review reports in general. On balance I think we will have a good review because we have good people on our panels who are working very conscientiously (which was the basic premise of the Beach team report) and because we have equally commendable OE staff and leadership. I would only caution that the difficulties and problems are

a/ See also a related document, "Comments on the Rite of Passage Review in Relation to the 'Negative Goals' of the Beach Team Report."

very real and that it would be reasonable to expect that it will take several years of hard work and iterative modification before the system can become exemplary. In the meantime we should be proud of what has been accomplished under the circumstances but modest about absolute achievement.

The remarks are organized around key features of the design, steps that were taken to operationalize the feature, and observations concerning what appear to be the results to date.

A. Specialist Panel Personnel. The "Beach Team" plan which provides the framework on which this review is planned stresses the importance of engaging top flight reviewers.

1. This general objective can be elaborated into three sub-objectives:
 - a. Choose high caliber people
 - b. Get representation from a wide range of relevant disciplines and professions, plus minority representation
 - c. Select the appropriate mix for each panel
2. The specific methods and procedures used to obtain nominations and make selections are not generally known; they should be written down for the benefit of those who will be continuing to work with the program. What glimpses I have had of it suggest that a broad

net was cast and that some outstanding people both gave nominations and were nominated. However, there may have been problems.

- a. Were data collected systematically which recorded who made the nomination? How many times he was nominated? What strengths and weaknesses were mentioned in making the nomination? Does any record exist except a file of vitas?
- b. Because of the lateness with which people were asked to participate and resultant schedule conflicts, many were called but fewer were chosen. It would be worthwhile to make a tabulation comparing those who finally served with those who were unable to serve. Was any kind of "bias" introduced at this stage? (e.g., higher caliber people found it harder to clear their calendars?)
- c. Some good work was done in identifying the types of talents needed on each panel. Again, a tabulation should be made to determine whether the final selection met these specifications. Just what disciplines and skill areas are represented?

B. Size of Panels. At an early stage it was expected that all panels would have approximately five members. At the last moment panel size was greatly increased, and actually ranges from seven to thirteen. The reasons for this change have never been stated, but the use of large panels can be expected to have a number of consequences.

1. Each of the clusters is certainly more homogeneous than the total set of programs. Nevertheless, each still covers a fairly wide domain, and an analysis revealed that each required a review panel representing a wide range of skills and backgrounds. This was probably the main reason for enlarging the panels.
2. When panels grow in size the group dynamics change. Large groups are more subject to the influence of dominating personalities. Also it is difficult to maintain the rule that everyone evaluates everything; the pressure for a division of labor grows. Meetings become very difficult to arrange with so many calendars to clear, and so the easy way out is to minimize site visits and meetings.
3. For the future it is suggested that panels be reduced to the original size of about five. This will be practical if the number of panels is increased, thereby reducing the number of programs assigned to each panel. In this way the programs will be more similar and the range of skills required correspondingly narrower.

C. Orientation of Panel Members and Chairmen. A fair amount of effort went into orientation activities. Chairmen were oriented separately and in advance of the three day orientation sessions for panel members. Packets of written materials were supplied in advance, although no BPP's were seen prior to the three-day sessions. A few individuals had prior experience as advisers or evaluators for the Lab and Center Programs, and others had contact with one or more units in other roles. However, the orientation of panel members (apart from chairmen) was not as thorough as would have been desirable.

1. It is doubtful that more than a few individuals have an understanding of the objectives of the program, its history, and the nature and significance of the change in support policy and assessment procedures. Although the Frye paper was distributed, it received little discussion.
2. In the 4-H orientation session, involving five of the panels, the orientation session was skipped.
3. Generally the orientation sessions were too open ended. OE should have done far more home work in advance. For example, the "Reports to be Generated" document could have been written just as well before the first orientation meeting; even though it was available at the beginning of the second meeting, it was not distributed until the afternoon of the second day. Similarly, most rules (e.g., site visits) could have been laid down in advance.
4. Many important design features for the review were left undetermined; in fact some have not yet been resolved. For example, the legitimacy of a panel reviewing and questioning the objectives of a continuing program is not clear. (The Beach Team report suggests it is not legitimate.) Also, everyone seems to say that cost/benefit analysis is important, but most of the panels seem to have declared themselves incapable of judging cost elements.
5. The role of the Master Panel, and particularly the kinds of decision alternatives it would be making recommendations for, were not clearly defined and differentiated from the role of the Specialist Panels. The attempt to say that SP's were obtaining

information and not evaluating was only confusing. Clearly the SP's must evaluate; but they are not making the same evaluations the MP will be asked to make.

6. At one time a rule was enunciated that no one could be a panel member who did not attend the three day orientation meetings. Some attended for only one or two days, and a few were actually appointed after the orientation meetings. The understanding of these people of the total program and of the review is likely to be deficient.

D. Cross-Institutional Substantive Review of Programs. In many ways this is the most promising aspect of the new design. Program managers in DRDR had already experimented with it in a limited way and were committed to moving farther in this direction.

1. Two key advantages are anticipated:
 - a. Comparative judgments can be made among substantively similar programs by the same reviewers
 - b. Review criteria and comparative analyses can focus on the substantive issues of the problem area rather than be limited to the process/product kinds of considerations that are common to all problem areas
 2. At mid course assessment the advances which can be anticipated in these two fields are very modest.
-

- a. Comparative judgments can and will be made. Having the same people judge all the programs to be compared is part of the requirement here, and this is being adhered to in large part although there are continued pressures from the panels to institute a division of labor. It appears in retrospect that rather than having large panels each look at a large number of programs it would have been better to have a larger number of clusters and smaller panels. In this way the clusters would have been more homogeneous, the range of reviewer skills needed would have been narrower, and the logistics problem would have been reduced. Another problem may still be susceptible to correction. The orientation process and criteria development focused almost exclusively on the review of individual programs. The fact that individual specialists would be asked for a ranking of programs or that the panel would be responsible for a comparative analysis emerged only at the end of the orientation period. Finally, the decision to insist on a simple rank ordering of programs seems unfortunate. It invites later "meat axe" decision-making. It is urged that the ranking be supplemented by a profile analysis of program which will show the ratings of each program within a cluster on individual criteria or criteria sets. Further, the importance of the panel's "cluster report" should be re-emphasized.

- b. It remains problematic whether the movement toward a more substantive level of analysis will take place to a meaningful degree. The original concept of general criteria and cluster criteria was a good one, but all panels seem to have either rejected the concept of cluster specific criteria or have adopted variants of the general criteria without moving to the cluster level. Indeed, NCERD leadership has now dropped the distinction between general and cluster criteria. With few exceptions (i.e., Panel E's introduction of criteria related to target populations) the criteria seem to have no special relationship to the nature of the problem area. It thus appears that one of the chief advantages of the new system will be lost unless it can be retrieved by emphasizing several parts of the panel's cluster report, namely the statement concerning the cluster domain and the analysis of similarities and differences derived from this conceptualization.
- c. When classifying complex multi-faceted programs such as those contained in this review it is to be expected that a program assigned to one panel will also have elements which are closely related to programs assigned to another panel. This will occur regardless of the classification system used. In several cases one panel has asked other panel or some of the members of another panel for an advisory opinion on a program. This is a useful practice and should be facilitated.

E. Review Methodology. In an effort to improve the reliability of the review process a number of procedures and rules have been adopted. Each of these design features can be judged with respect to its intrinsic validity, the degree to which it is being implemented, and what appear to be the results so far.

1. The Anti-Site Visit Syndrome: There is a set of rules based on a distrust of personal interaction between reviewers and unit personnel.

a. They include:

- 1) The number of visits to demonstration sites is limited to one or two per panel.
- 2) Units may not be visited.
- 3) Only one (possibly two) representative of the unit may interact with the panel in person, either at the demonstration site or in Washington.

b. It is not clear why these rules have been instituted. The Beach Team displays an attitude which is skeptical about interaction with unit personnel, but makes it clear that Specialist Panels (but not the Master Panel) are expected to make site visits. Apparently there is a fear that panelists will be overly influenced by the persuasive charms of unit personnel. Without supposing that this has never happened, I maintain that this view is short-sighted and ignores several problems created by the new rule.

- 1) It sells the intelligence and perceptiveness of our panel members short.

- 2) Personal interaction is more likely the means for detecting shortcomings than providing the means for undue influence.
 - 3) Written documents are just as likely to be "snow jobs" as personal presentations.
- c. Similarly, limitation of interaction to one or two representatives is difficult to understand. We are not reviewing one-man projects. The programs being reviewed are highly complex large scale undertakings requiring teamwork of multi-disciplinary groups. On the panel side this complexity has been recognized by assembling a multi-disciplinary panel. The idea that one individual should represent the entire program is a contradiction of the basic assumptions underlying the establishment of the Laboratory and Center Programs.
- d. Finally, choice of demonstration sites in preference to the unit site is a questionable choice.
- 1) What one can see at a demonstration site is such a small sample of the program and the people working with it that reliability is very low.
 - 2) Such a visit necessarily focuses on past performance and products. While such information is useful, the means of making the inferential jump from the past to judgments about a plan for the future are not always clear. It thus serves to obscure the fact that each panel is being asked to make judgments about a plan for the future.

e. On the other side of the coin, the chance to interact with members of the program team along with representatives of management and support services has many advantages.

- 1) One can judge the depth and variety of talent necessary to fulfill the plan.
- 2) Questions can be asked that dig behind the written word.
- 3) Perceptions and impressions can be tested for validity and corrected if necessary.

f. For the first month of the review these rules seem to have been the source of considerable confusion, and it was not until May 9 that any specific (demonstration) site visit was approved. At this point it appears that not many visits will be made, but this may change.

g. It is doubtful that a document oriented review will have full face validity with Congressmen and others who will be judging this review. In particular it will be very hard to make an adverse judgment stick if the program has never been visited or the program staff questioned. In recognition of this situation NCERD leadership has now instituted a fail-safe strategy for down-rated programs. Although the timing has not been decided on, at some point those programs which are likely to be given a negative rating will be identified, and ~~arrangements will then be made to make a site visit to the unit in~~ order to verify the preliminary perceptions. This should reduce the

number of "Type A Errors", i.e., making a negative judgment about a good program. Unfortunately there is no comparable safeguard against making a "Type B Error", i.e., making a positive judgment about a poor program.

2. The 100% Participation Rule. One of the key guidelines for the review is that all panel members should review and evaluate all programs assigned to the panel, and that all, or as close to all as is feasible, should make any visits that are undertaken. This seems like a good rule designed to increase the reliability of judgments. However, the matter is more complex than it may appear, and there are real problems of feasibility with respect to site visits.

- a. Clearly if different people rate different things, there is a problem of comparability of the judgments. One ordinarily gets all judges to rate the different things in order to get around this problem; the judges use the same criteria, and one can even correct for tendencies by some individuals to rate high and others to rate low. However, in the classic case, the judges are deemed to be equally competent. That assumption may not apply to the present situation. The panels were deliberately selected so as to represent the wide range of skills necessary to judge the complex programs under review. In this situation each judge is not equally competent to review every aspect of every program.

To force them to do so may introduce an element of spurious reliability (panel members not expert in one element duplicate the ratings of those more competent in it--a halo effect) or reduce the validity (the non-expert judge goes ahead and makes an independent incorrect evaluation). It may be difficult to avoid this problem within the range of feasible panel size and the constraints of cost, but it does not follow that the 100% rule should be followed rigidly. It is suggested that panel chairmen be allowed to use this discretion concerning the introduction of at least some degree of division of labor. A number of panels are already doing this to some degree, and it should be legitimized.

- b. With respect to site visits there is also a problem of the feasibility of enforcing the 100% participation rule. Getting 10 to 13 busy people to agree on a date is very difficult. The chief result of the rule seems to have been to discourage panels from making any visits. The loss of information may not be worth the gain in consistency.

F. Problems of Consistency Within and Across Panels. An important objective is to have all panels operate on the basis of common rules and procedures. This is not to say that activities must be rigidly uniform at all levels of specificity; there must be some room to respond to the special needs of specific program and situations. But at some level we must be able to say that a common methodology was used. Lack of consistency occurs in several areas.

1. Some programs will be reviewed at demonstration sites while others will not.
2. Panels have interviewed unit personnel of some programs in Washington, but not of others. (One panel adopted this as a uniform procedure.)
3. The rule of one representative has been progressively modified.
4. Some panels have adopted a division of labor for "in depth" reviews, while adhering to the rule that all members will review all programs; others have not.
5. As mentioned elsewhere, the general criteria are similar without being uniform.
6. In at least one case a suitable majority of panel members went into the field, but they split up to visit six different sites.